Analyzing Healthcare Big Data for Patient Satisfaction

KaiYu Wan
Department of Computer Science
and Software Engineering
Xi'an Jiaotong Liverpool University
Suzhou, China
Email: kaiyu.wan@xjtlu.edu.cn

Vangalur Alagar
Department of Computer Science
and Software Engineering
Concordia University
Montreal, Canada
Email: alagar@cs.concordia.ca

***Abstract*—Healthcare Big Data (HBD) is more complex than Big Data (BD) arising from any other critical sector because a variety of data sources and procedures are followed in traditional hospital settings and in healthcare network (e-Health). In order to achieve their primary goal, which is to enhance patient experience while sustaining dependable care within financial viability and respect for government regulations, the HBD should be analyzed to determine patent satisfaction level. In general, there exists no accepted method yet in measuring patient satisfaction. The traditional approach for evaluating hospital-based healthcare is through a statistical analysis of responses of clients to a survey, often conducted by a third party. Such methods are often infected with incomplete information, inaccurate hypothesis, and error-prone analysis. Analyzing data generated through automated healthcare networks for assessing the effectiveness of service provision and patient satisfaction are more challenging. It is in this context that we discuss in this paper factors that contribute to patient satisfaction, and propose an algorithmic method to assess it from HBD analysis.**

Keywords: Big Data, Health Care Domain, Hospital-based Services, e-Health Services, Patient Satisfaction Analysis

I. Introduction

Historically, the healthcare industry is one of the largest generators of data. The term "Big Data (BD)", although first used by Cox and Ellsworth in [1], was conceptualized by Laney [2] with the keywords "Volume, Variety, and Velocity". These keywords characterize healthcare data, such as patient records, employee records, and clinical information, that hospitals generated many decades before BD concept was conceptualized. Of late, healthcare data has grown massively in both structured and unstructured form, largely driven by the demands of constantly expanding knowledge-thirsty population and operational characteristics of e-Health platforms. This explosive multi-dimensional growth has lead researchers [3], [4] to add many more keywords to characterize Healthcare Big Data (HBD). In this paper we restrict to the 6 Vs "Volume, Variety, Velocity, Value, Veracity, Validity". It is known [5] that in US alone the volume of healthcare sector data had reached 150 exabytes (10^{18}) in 2010, while the volume of healthcare data in China and India are approaching 10^{24} (yottabyte). It is not just the volume but their variety, namely the types of sources that produce data and the target types that demand them are too diverse and numerous in Healthcare domain. These include healthcare personnel (physicians, clinical staff, care givers), service providing agencies (including insurers), hospitals with assets, psychologists, government regulators, pharmacies, pharmaceutical manufactures (with research teams included), and medical device companies. Online healthcare monitoring and time constrained clinical diagnosis need to be done while data is streaming through the network. Velocity refers to this aspect. Value refers to the healthcare specific valuable determinants that are necessary to achieve the goals set by healthcare institutions. Veracity in HBD refers to the reliability and truthfulness of data sources, and Validity refers to the correctness of data in every context of usage. Any inaccuracy or uncertainty in the data from any of its support sector assets will affect the trustworthiness of the healthcare sector. As an example, a non-certified medical device or a virus attack on an asset of the communications sector might disrupt the availability of medical services to chronically ill elders. Consequently, HBD is more richer than BD generated in other sectors and is difficult to comprehend, model, manage, and analyze.

A. Contribution

In [6] we investigated different sources of HBD and their types, and suggested a hierarchical classification of HBD. This classification is intended to suggest different methods for analyzing HBD that belong to different groups in the classification. Several analysis methods that are currently being researched and tool support for such analyses are discussed in [3], [4], [7]. Their goals differ from our goal in this paper, which is "analyzing HBD for patient satisfaction". In Section II we briefly discuss the impact of the 6V's that we have chosen to characterize HBD on patient satisfaction. We motivate in Section III the notion of "patient satisfaction" as an emergent property from *Performance*, *Process*, and *Purpose* (denoted P^3) that govern healthcare services, whether they are provided straight out of hospitals (TRD) or e-Health platforms (EHD). Based on this hypothesis we discuss methods in Section IV for analyzing data generated from TRD, and data generated from EHD. We conclude the paper in Section V with a justification for our approach.

978-1-5386-2165-3/17/$31.00 ©2017 IEEE

II. Characteristics of Healthcare Big Data

In this section we explain the meanings of the 6 Vs that characterize HBD that are generated in TRD and EHD paradigms.

- *Volume:* It refers to the *massive amount* of structured and unstructured data generated by various healthcare agents. Volume also refers to the rate at which the amount of stored/processing data grows. The trend is that in TRD record keeping for every event is necessary for regulatory and compliance requirements. Patient information is *accumulated* every time the patient walks in to the hospital, and every patient-physician conversations are recorded. With wearable devices and mobile health platforms there is a continuous update of patient's health profile in EHD platform. Many real-time imaging technologies and point of care devices are required for providing "Personalized Healthcare Services". In this venture researchers are studying a "Virtual Physiological Human" (VPH) framework that will enable a collaborative investigation of human body as a single complex system [4]. In this study massive amount of data is expected to be generated from *genomic* (the branch of molecular biology concerned with the structure, function, evolution, and mapping of genomes), *proteomics* [8] (the large scale studies of protein), *metabolomics* [9] (the intermediates and products of metabolism). In addition, data on the environmental attributes that are necessary to monitor and predict personalized health must be collected by continuous sensing [10]. The volume feature arising from the health service sector has an impact on data analysis, in particular in devising efficient methods to assess patient satisfaction.
- *Velocity:* It refers to fast streaming of data in eHealth environment. New technologies for healthcare monitoring that use sensor-based medical devices for continuous monitoring of patients generate data that is in motion. The frequency and the speed with which data is generated, periodically or asynchronously with other observations, must be checked for correctness and consistency before being consumed by clinical tests. If the process of collecting and utilizing data does not meet the speed with which data streams through there is a potential for data loss or data corruption. This will adversely affect patient satisfaction.
- *Variety:* This refers to *heterogeneity* of data. The three aspects of heterogeneity arise from *different data types* and *different sources (schemas)*, and *different data functionality*. In fact, a source produces one type of data. Thus, variety describes different data formats that do not lend themselves to storage in structured relational database systems. Typically, HBD includes structured data (Electronic Health Records), semi-structured data (graphs, XML files), and image files (MRI, CAT). They have schema heterogeneity. Frequently, clinical data and prescriptions are represented in natural language. Each kind of data will be queried in a certain manner that requires a specific kind of "operation". The correctness and efficiency of query processing impacts the quality of health care service rendered to patients,
- *Veracity:* This refers to reliability and truthfulness of data. In EBD sensor data may be inaccurate, and errors arise due to faulty medical devices and their observations in non-standard format on humans. Data streaming in from sensors attached to human body and data coming from mobile devices used by healthcare actors have different speeds. Synthesizing them in making correct decisions are error prone. Data sources must be reliable, data communication must be reliable, and recording them must be truthful. The use of undamaged data from social media, and personal health data entered manually by patients on-line are likely to introduce errors in analysis and prediction. In TRD it is common that medical personnel or care givers unintentionally enter inaccurate information in the health record of patients. Both volume and velocity have a combined impact on veracity. It is hard to prevent inaccurate measurements that arise from devices and software-enabled data transmission. Hence, decision making must be based on robust algorithms that work with uncertainty in data measurement. There is also a human factor that impacts veracity in TRD, which is the "attitude/care taken by medical professionals" in entering the data correctly. Due to the extreme pressure that medical professionals face in the work environment, they see the "logging of data a bureaucratic need and waste of time that distracts them from the healthcare care service they provide to patients [4]. Consequently, clinical practice tends to generate low quality HBD, while HBD generated in clinical research are of rich quality. This aspect has both positive and negative bearing on patient satisfaction.
- *Value:* Value refers to the basic "business values" associated with healthcare domain. In particulate, this involves the achievement in *patient satisfaction*, *affordability and cost optimization*, and *develop, expand and improve EHD*. Even if one of them is not achieved the potential fall in "economic value" is very high. Many research organizations who buy healthcare data from hospitals and clinics may choose to store HBD even before charting a plan for using them to gain improve its business base. The effort to store and analyze HBD must be based on "value" that meeting its mission.
- *Validity:* This characteristics refers to the "correctness and accuracy" of data. After data is gathered from different sources it should be validated with respect to "expected" properties. For example, validity of patient personal information such as age and blood type must be performed before treatment commences. Data generated during seasonal health effects and disease evolution should be subjected to statistical tests in order to remove noise and filter out inaccuracy in order to meet domain criteria. Data collected from the same population during

typhoon and floods, earthquake, and chemical leaks will have "a different flavour" from the same data collected at normal situations. That is, data could be valid in one context but not valid in another context. Patient safety is affected when data is not validated.

III. Patient Satisfaction

The main goal in healthcare industry is to deliver *reliable* service which is perceived by patients to be *satisfying*. Patient satisfaction determines the loyalty level of patients [11], and consequently is directly related to economic viability and stainability of the Healthcare Service Provider(HSP). In this section we discuss the key features of healthcare service that drives reliability, and the antecedents to overall patient satisfaction.

Every patient creates a personal health profile using the spiral model [12], and makes it available to the HSP from whom healthcare services are sought. This "patient-centric" healthcare data model serves as a basis to assess whether or not the patient receives relevant healthcare services. The data in this model are very specific to the health status, living conditions, and the social network of the patient. This *Personal Health Model* (PHM) of a patient is enriched with patient-specific privacy rules and healthcare-centric ethical rules. The PHMs of a patient are given to her HSP who will guarantee their confidentiality. Informally we may say that a patient is *satisfied* if the services delivered by the HSP *meet the goals and situations* as specified in her PHMs.

A. TRD Platform

Patients who visit hospitals for receiving healthcare services are directly confronted with (1) administrative procedures, (2) infrastructure-based facilities, (3)interaction with clinical staff and clinical test procedures, (4)interaction with nursing staff, and (5) interaction with physicians. Zeithmal [13] has argued that patients who cannot judge the technical quality of the service would base their quality judgements on structure and process dimensions such as physical infrastructure settings (waiting area appearance, quality of restaurant), the ability of staff to solve their problems (courtesy, empathy), and waiting times. Hence, patient satisfaction which is a mixture of emotions, the tangibles, and intangibles cannot be viewed in the same light as that of "consumer product satisfaction". While patient satisfaction has become the norm to determine "value of service delivered", it is regarded as the most difficult measure to quantify and measure.

Many studies reviewed in [14], [15] have shown that there is a strong relationship between *employee satisfaction* and *patient satisfaction*. They agree that measuring satisfaction is a complex process because it takes into account *expectations*, *outcomes*, *perceptions*, and *personal evaluation*. These in turn can be categorized into *Employee Satisfaction* (ES), *Employee Engagement* (EE), *Employee Burnout* (EB), and *Job Satisfaction* (JS). Below we give a brief summary of factors, selected from the published results of Gibbons [16], Kahn [17], Janicijevic et al. [18], Ott and van Dijk [19], and Harter et al. [20], that contribute to each one of these categories.

[ES:] The factors that contribute to this category are adequacy of work resources, interpersonal relationship, opportunities for advancement, time allocated for accomplishing a task, salary, working environment, and clarity of instruction on job expectation.

[EE:] The primary factors that contribute to this category are (1) employee's ability to interact with patients and keep them engaged in their daily heath-related activities (taking medication, personal hygiene, exercise), (2) employee's relationship with senior supervisors in the organization, (3) job characteristics and their compatibility to the skill and knowledge of the employee, (4) rewards and recognition of loyalty and ethical behavior, and (5) procedural justice/flaws.

[EB:] The primary reasons are (1) the inability of an employee (nurse, physician, lab technician) to complete the allocated job within predefined time limits, (2) insufficient resources to meet a large demand from patients, and (3) larger than expected number of patients to care in a day.

[JS:] The primary factors are (1) meaningfulness of job, (2) safety of work place, (3) reward and recognition, and (4) emotional/psychological willingness to engage with the tasks.

Many factors arising from infrastructure facilities and administrative procedures affect patient satisfaction. These include the physical or cyber infrastructure elements that create experience and affect emotions. These include (1) ambience of the waiting rooms, (2) waiting times for admission and release, (3) hygiene in cafeteria, waiting halls, and toilets, and (4) accessibility and safety inside and outside the hospital area. These infrastructure factors may interfere with some of the employee satisfaction factors, influencing patient satisfaction. For example, an employee who does not like cafeteria facility for lunch because of the long waiting line, sparse sitting/eating area, and quality of food might be less engaged (EE factor) and might suffer burnout (because of the degree of chaos experienced in the cafeteria).

B. EHD Platform

EHD platform is an open patient-centric distributed service system in which patients, physicians, medical devices, and clinical units are networked, and are empowered to share knowledge. In this platform there is no *direct* interaction between patients and healthcare employees. So, we hypothesize that patients are satisfied only if the healthcare system is *trustworthy*, in the sense that services are reliable and are provided to patients at all times regardless of where they live. Once patients *trust* the system performance and quality of services delivered by it they become loyal customers of the system. In this platform health service utilization is maximized, service cost is minimized, universal health coverage, and universal health awareness is imparted to patients. Thus,

there is potential to achieve a high level of level of patient satisfaction, once patients trust the healthcare system and hence the HSP behind the system's activities. A patient's trust of EHD is a combination of *trusting* the HSP, and *trusting the automated system* which is the medium for providing heath services. We briefly explain the factors that contribute to trusting a service automation system [21] and relate them to "trusting behavior of patients" in a healthcare automation system.

The trust in HSP is a combination of *cognitive trust* (CT), which is consumer's confidence and willingness to rely on the *competence* and *reliability* of the HSP, and *affective trust* (AT), which is characterized by the confidence placed on HSP from "emotions" generated by the level of concern demonstrated by the HSP. The four-fold important determinants of CT are the following.

1) *Expertise of HSP:* This will convince the patient on the competence, understanding, faith, and motives of the HSP.
2) *Service Performance:* This informs the patient in advance the set of *quality attributes* of healthcare service provided by the HSP.
3) *Similarity:* This aspect will emphasize the *commonalities* between the goals of patient (getting services) and HSP (delivering services).
4) *Reputation:* The service rating of HSP assigned by the Medical Board Authority (MBA) must be made public. Affective trust (AT) is built from *Reputation*. The HSP, through the on-line system, might also advertise "promotional items" and introduce its trusted clients who have formed social groups. Consumers who participate in this "socialization" process might be persuaded to get into trusting relationship with HSP.

A patient's trust in the on-line healthcare (automated) system is determined by "the *belief* that the system will help her achieve the health goals (as specified in the PHM), even in situations characterized by uncertainties and vulnerabilities". Disconfirmation theory [22] has been accepted as the most dominant conceptual model for *consumer satisfaction*. It states that "consumer compares her perception of the service against a pre-purchase expectation". It is known [14] that many researchers in medical field that "marketing-oriented consumer models" do not fit well with "patient-oriented medical scenarios", and hence we may conclude that "consumer satisfaction, as in e-Commerce" is very different from "patient satisfaction in on-line healthcare systems". After a careful study of patient-employee interaction scenarios [22], [23], [14] and human-system interaction in "service automation scenarios" [24] we have put together a mix of several dominant determinants for assessing patient satisfaction in healthcare automation. This is discussed next.

Trust of consumer in automation was classified [24] into one of three dimensions *Performance*, *Process*, and *Purpose* (PPP). We can specialize them for healthcare service automation as follows.

Performance: It is to be regarded as a combination of *Competence*, *Information Accuracy*, *Reliability*, and *Accountability*. Competence is patient's estimate of the suitability of the healthcare system to achieve her health goal, as indicated in her PHM. Information accuracy is patient's judgment on the level of precision of the healthcare information given by the system. Reliability over time is patient's prediction on how long (and how far) in the future the current health information given by the system will remain valid. Responsibility reflects patient's assessment on the functional completeness of the health information system.
Process: It is to be regarded as a composition of Dependability, Understandability, Control, and Predictability. Dependability is patient's assessment on the consistency and reliability of system response to her interactions. Understandability reflects the patient's ability to grasp the functionality of the healthcare system. Control stands for the patient's emotion "am I in control of my personal information or not?". Predictability is patient's assessment of "whether or not her expectations are met" in receiving healthcare.
Purpose: It is defined to include Motives, Benevolence, and Faith. Motives assess "whether or not the intended use of the system is truthfully communicated to patients". Benevolence is patient's view on "whether or not the system has her interests and goals while delivering services". Faith is the belief that the system can be relied upon in future care.

With the above factors we add the antecedents for patient-level trust. We adapt the three economic theories [25] that was proposed for e-Commerce. These are respectively called *Transaction Cost Economics* (TCE), *Information Economics* (INE), and *Socio-Economics* (SOE). We explain how this may be adapted for improving patient satisfaction in healthcare sector.
Transaction Cost Economics (TCE): It states that "every service transaction has a quantifiable direct and indirect cost, as well as an unquantifiable cost". In healthcare sector, the direct cost is the cost involved in patient monitoring, diagnosis, clinical testing, and healthcare service implementation. The indirect cost may include cost of healthcare information dissemination through healthcare network and other media, and monitoring the different transaction phases. The unquantifiable cost is usually the "cost of disadvantage" in the event of posting inaccurate or misleading or incomplete information, contract violation, and not meeting service provision rules within certain time constraints. This TCE theory may be adapted by HSP in the following manner.

- Information on healthcare services should be disseminated to patients *without interruption in a proactive manner*. Human interaction and processing associated information in a timely fashion are central to patient-care. So, information processing and dissemination must be *optimized*. The HSP must increase the knowledge and level of understanding of patients, and improve on-line protocols for patient-physician interaction. All information collected by the system in every interaction

scenario must be correctly recorded and securely saved. Consequently, *Volume*, *Velocity*, *Veracity*, and *Validity* characteristics of HBD greatly influence the level of patient trust. This has implications to .

- The HSP should explain to patients how processing complexity will be contained, because complexity causes errors. The patient should be made to understand the medical treatments that are suitable for her health status as specified in her PHM. For each treatment the process for getting it should be made clear to the patient. The patients should be educated on drug doses, times of administration, correct methods for using medical devices, different methods of surgical dressings, and post-treatment options. The information communicated by HSP to patients must be *correct and current*. That is, HSP must ensure integrity of information in spite of its *Velocity* and *Variability*. Patients must be informed on security and privacy aspects for sharing their medical information in the healthcare network, and told the conditions for delivering services in total confidence. Hence, *Veracity*, *Value*, *Variety*, characteristics of HBD also influence the patient trust.
- HBD *Volume* complexity must be reduced by using the PHMs of a patient as *filters*. That is, only health information that is *relevant* to a PHM will be delivered. Using the same filtering (based on context information extracted from a PHM) services whose quality violate the ethical and social value in an environment should not be provided by the HSP. That is, safety, respect for social and cultural norms, and ethical principles (especially for servicing elderly and physically challenged) must be explicitly respected.

Information Economics Theory (INE): The INE theory should be adapted for healthcare to overcome the *asymmetric* relationship between the patient and HSP. In general, HSP is more knowledgeable about its services than a patient, who may have little or no knowledge on what services to expect from a HSP. Since the patient is most likely to accept only services that have potential benefits for her and high economic value (low cost), the HSPs should provide knowledge-intensive tutoring to the HCR before delivering health services. It is essential that the web site oh HSP is "patient-centric", enabling them to "browse" and search for "relevant" medical information using hand-held/mobile devices. A patient should be convinced of the authenticity of information provided, and understand clearly the procedures for selecting services, completing the contract, and stating their preferences for service delivery. Intelligent patient-centric user interfaces in which PHMs are used as filters and authenticators would enable relevant and accurate health knowledge delivered *directly* to patients. The *learning curve* required by patients to understand the system functionalities is thus made "less steep" through enabling context-dependent "long term" tutorials. Consequently, the HSP can "build trust through direct interaction with patients". Since the PHMs of individuals are directly uploaded to the system, it becomes easier for HSC to build automated mechanisms in building trust with patients.

Socio-Economics Theory (SOE): We adapt the SOE theory to healthcare in suggesting that the HSP pays attention to "trust recommendations" and include them in their service publication. Social networks, family and cultural links, and environmental determinants, specified in the PHM of a patient, will influence patient behavior. Patients may seek recommendations from their families, and social groups in choosing the physicians, labs, and nursing homes. They may consult other trusted healthcare authorities (such as government reports) to guide them in choosing best healthcare providers. Hence the HSP should willingly submit to independent ranking authorities and patient groups for their evaluation (based on questionnaires).

IV. Patient Satisfaction Assessment

Since July 1, 2011 patient satisfaction and service quality scores have become the most important criteria for Medicare reimbursement [14]. Hence there is a flurry of research activity in choosing the relevant determinants to measure patient satisfaction and create a model of it for assessing patient satisfaction measure. Many measuring instrumentation methods are being studied [26] to assess patient satisfaction with healthcare services, satisfaction with inpatient satisfaction, satisfaction with out-of-hours primary care, and satisfaction with postoperative surgical care. Thus, patient satisfaction is multi-dimensional. Many works [14], [27], [26] have focused only on single-dimensional associations, such as "the impact of hospital work climate on patient satisfaction" and "patient's overall satisfaction with primary care physicians". Their approach is the traditional questionnaire-based statistical factor and regression analysis. Typically, a questionnaire includes a finite number of items, where each item is chosen by the investigators after expert's confirmation of its relevance. For example, "My doctor examines me carefully", "My doctor gives me enough information before ordering a test", and "My doctor shows respect to what I have to say" are relevant items because they are directly related to patient's satisfaction with the physician. Each item is to be answered by a patient on a 5-point Likert scale (1= completely dissatisfied, 2=somewhat dissatisfied, 3=neutral, 4=somewhat satisfied, 5=completely satisfied). Total satisfaction score is calculated from sum of response from all items in the questionnaire. The standard statistical methods, such as "Factor analysis", "Pearson product-moment correlation coefficient", "Regression analysis" [28] are used to examine the strength of causality, relationships between variables, and group differences. Some recent works have explored a multi-dimensional analysis for patient satisfaction. In [27] the authors study patient satisfaction and trust as factors driving "loyalty to Primary Care Physicians". Based upon the questionnaire-based method and analysis they have concluded that (1) there is a strong positive relationship between patient satisfaction and patient loyalty, and (2) patient trust (includes trusting relationship with physicians) fosters deep loyalty". Healthcare service quality was consid-

TABLE I: A Sample of ERS Ratings

Questionnaire	[5]	[4]	[3]	[2]	[1]
1. Overall Rating	32	132	64	36	36
2. Level of Care	96	34	130	32	8
3. Wait time	90	48	64	34	64
4. Information Provided	50	146	60	32	12
5. Facility	60	130	96	6	8
6. Physician	60	156	60	32	32
7. Nursing Staff	76	88	90	40	6

ered in [29] to be a vital determinant of patient satisfaction and patient trust. They remark "the causality relationship between patient satisfaction and patient trust does not appear to be very clear in the literature. Depending on the study, satisfaction appears either as an antecedent or as a result or consequence of trust." In our study, we have posited that trust is essential (antecedent) for satisfaction. The study [30] is restricted to explore the antecedents of patient trust in healthcare insurers. Their conclusion is that "satisfaction with insurers and trust in physicians are strongest predictors" of trust in health-care insures. The analyses methods for "multi-dimensional" setting is to create many questionnaires, one specific to each dimension, create tables of responses, and analyze them for interrelationships. Two things are clear from this review. One is that "no standardize procedure exists that comprehensively measures patient satisfaction" and "several hypotheses have been put forth whose statistical analyses have not given definitive answers on their validity". Our model, based upon the antecedents discussed in Section III is multi-dimensional and has three steps. In the first step we assess patient satisfaction from response collected from patients. This gives the "patient-centric measure". In the second step we assess patient satisfaction from responses gathered from medical experts. This gives "expert-centric" measure. In the third step we use these two measures, expressed as vectors, to calculate "the cosine measure" of the two vectors. We use Schulze's method [31], [32], explained next, in the first two steps.

A. Analysis

Let $\Lambda = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be the set of antecedents (determinants) of patient satisfaction. An example of such a set is {Emergency Room Services, Admissions & Discharge services, Physician Services, Nursing services, clinical services, Ambulance Services, Nutritional services, Reliability & Availability of Staff}. The patient satisfaction for each α_i is determined, using a statistical analysis on the one-dimensional questionnaire-response model. Let N be the number of patients whose views on the relative importance of α_i, $1 \leq n$, will be taken into account for assessing patient satisfaction. We use the standard questionnaire model [33], [34], of HCAHPS (the Hospital Consumer Assessment of Healthcare Providers and Systems) which is a patient satisfaction survey required by CMS (the Centers for Medicare and Medicaid Services) for all hospitals in the United States. The Survey is for adult inpatients, excluding psychiatric patients. As a simple example, a questionnaire for Emergency Room Services (ERS) with 7 items and the responses from 300 patients is given in Table I. The columns are the 5-point Likert scale numbers, with [5] in column 1 being the highest level of satisfaction and [1] in column denoting the lowest level of satisfaction. The rows are the questions (in some random order) included in the questionnaire. Table entries denote the number of patients responding, for example 90 patients are "completely satisfied" with "wait time", 156 patients are "somewhat satisfied" with "physician", and 8 patients are "completely dissatisfied" with "Level of Care". The total of all entries in each row is 300, reflecting that all patients responded to the survey. For each α_i a table is constructed. All such tables have the same number of columns, although the number of rows will be different because the type and number of questions included in the questionnaire for α_is will be different. Instead of conducting a statistical analysis on each table, we introduce a new analysis technique that involves both patient-centric and medical expert-centric inputs on "preferences".

Step 1: User Preferences of Questions: We want to assign "user-defined preferences" to the queries in each questionnaire. The rationale is that in reality some patients might view "Level of Care" as *more important* than "Physician", whereas many others might view "Physician" as *more important* than "Level of Care". As such, the scores in the tables are neither "uniformly distributed" nor are they stable to a statistical analysis. We explain below how preferences are used in assessing patient satisfaction score. The preferences of users can be collected by asking patients who complete the questionnaire to "rank the questions". Since the user ranks will be different, in general there are N (=300 in our example) ranked lists of items. Using the method proposed by Schulze [31], [32] we can produce one ranked list, in which the ranked items will satisfy all patient preferences. This method is based on the definition of a Schwartz set, as used in the Schulze method: (1) An *unbeaten set* is a set of items of which none is beaten by any item outside that set. (2) An *innermost unbeaten set* is an unbeaten set that doesn't contain a smaller unbeaten set. (3) The *Schwartz set* is the set of items who are in innermost unbeaten sets. A brief outline of this procedure is given below.

Procedure SCHULZE [31], [32]

Suppose $Num[q_i, q_j]$ is the number of patients who strictly consider question q_i as more important than question q_j. A *path* from q_i to q_j of strength p is a sequence of questions $q(1), ..., q(k)$ with the following five properties: (1) $q(1)$ is identical to q_i. (2) $q(k)$ is identical to q_j. (3) For all $t = 1 \ldots (k-1)$: $Num[q(t), q(t+1)] > Num[q(t+1), q(t)]$. (4) For all $t = 1 \ldots (k-1)$: $Num[q(t), q(t+1)] \geq p$. (5) $p[q_i, q_j]$, the strength of the strongest path from question q_i to q_j, is the maximum value such that there is a path from query q_i to query q_j of that strength. If there is no path from query q_i to query q_j, then $p[q_i, q_j] := 0$. A query q is *preferred better* than query q' if and only if $Num[q, q'] > Num[q', q]$. Query q is a potential winner (among all queries) if and only if $Num[q, q'] \leq Num[q', q]$ for every other query q' in the questionnaire.

Remark 1: It is possible to prove that $Num[q, q'] > Num[q', q]$ and $Num[q', q''] > Num[q'', q']$, which implies $Num[q, q''] > Num[q'', q]$. Therefore, it is guaranteed (1) that the above definition of "better" really defines a transitive relation and (2) that there is always at least one query q with $Num[q, q']Num[q', q]$ for every other query q'.

The output of the above procedure gives an ordered sequence $\{q_1, \cdots q_m\}$ of the m queries in a questionnaire. We remark that q_1 is regarded "most important", and q_m is regarded least important by patients. We assign weights $w_1 \ldots w_m$, such that $0 \leq w_i \leq 1$, and $w_1 > w_2 \ldots > w_m$. The weights may be chosen arbitrarily subject to the above restriction. For the sake of stability in computing, it is better to normalize the weights. That is, $w_1 + w_2 \ldots + w_m = 1$. Also, reorder the Table of responses so that the rows are ordered according the query order (in decreasing order of user preference) $\{q_1, \ldots q_m\}$. This operation does not change user responses (as reflected in the Table). Now, we compute the five quantities $\eta_i = \sum_{j=1,m} w_j \times Table[j, i]$. The interpretation is that η_i represents the "contribution to patient satisfaction from the antecedent associated with the Table. As an example, assume that Procedure SCHULZE applied to user preferences on the 7 questions in Table I produces the query order $q_1 > q_6 > q_7 > q_4 > q_3 > q_5 > q_2$. We can choose weights $w_1 = 7, w_6 = 6, w_7 = 5, w_4 = 4, w_3 = 3, w_5 = 2, w_2 = 1$, and after normalization we get $w_1 = \frac{7}{28}, w_6 = \frac{6}{28} w_7 = \frac{5}{28} w_4 = \frac{4}{28} w_3 = \frac{3}{28} w_5 = \frac{2}{28} w_2 = \frac{1}{28}$. Thus, $\eta_1 = w_1 \times 32 + w_6 \times 60 + w_7 \times 76 + w_4 \times 50 + w_3 \times 90 + w_5 \times 60 + w_2 \times 96 = 58.7$. In a similar fashion we calculate $\eta_2 \ldots \eta_5$. We calculate the maximum of $\{\eta_1, \ldots \eta_5\}$. Let η_k , for $1 \leq k \leq 5$ be the maximum. We define η_k as the contributing factor for patient satisfaction from ERS. In general, for a Table with m queries that is associated with an antecedent, say X, η_k is calculated and defined as a measure of patient satisfaction contributed by the antecedent X.

Step 2: Calculation of Expert-centric Patient Satisfaction Score: We require the medical experts to give their *preferences* on the set of antecedents that are chosen to measure patient satisfaction. For example, if the antecedents to measure patient satisfaction are {Emergency Room Services, Admissions & Discharge services, Physician Services, Nursing services, clinical services, Ambulance Services, Nutritional services, Reliability & Availability of Staff}, then the medical experts (who might take input from healthcare managers) are better positioned to "rank" the antecedents in terms of their importance in assessing patient satisfaction. The rationale is that the medical expert's opinion of "importance of antecedent" should be factored into the final calculation of patient satisfaction. We do not allow any randomness in such assessments. In general, the set of antecedents $\Lambda = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is ranked by each expert. The rankings of the experts are merged by Schulze's algorithm (**Procedure SCHULZE**) and the output of this procedure is a ranked sequence of the antecedents that is not in conflict with any expert's opinion. For the sake of clarity let us denote this ranked output as $\{\alpha_1, \alpha_2, \cdot, \alpha_n\}$. We assign normalized weights θ_j, $j = 1, \ldots n$ (using the method discussed earlier) to this ordered list of antecedents. That is, θ_j is the weight assigned to antecedent α_j.

Step 3: Patient Satisfaction Measure: We calculate the "inner product" of the vectors $\{\eta_1^{k_1}, \ldots, \eta_n^{k_n}\}$ and $\{\theta_1, \ldots, \theta_n\}$, where $\eta_j^{k_j}$ is the patient-centric measure calculated in Step 1 for the j^{th} antecedent α_j, and θ_j is the weight assigned by experts to the j^{th} most important antecedent. The resulting measure is $\sum_{j=1 \ldots n} \theta_j \times \eta_j^{k_j}$. The interpretation of this normalized sum is that a value closer to 1 shows "maximum patient satisfaction" and a score below 0.5 will indicate the need to improve services to earn an acceptable level of patient satisfaction.

V. Conclusions

Patient satisfaction is a subjective evaluation of healthcare survives delivered to the patients. We hypothesize that once patients trust the service provider and the services both patient satisfaction and patient loyalty are bound to rise. In order to maintain a high-level of patient satisfaction service providers should improve the quality of services provided and convince the patients of the high quality of service, in order to earn their trust. After identifying a set of antecedents for patient satisfaction we followed the traditional procedure of "questionnaire-based survey method" to interact with patients. However, unlike almost all analysis methods that use standard statistical packages to analyze the patient responses, we factored in patient preferences (for questions in each questionnaire) and expert (medical and management) preferences in determining the set of antecedents on which questionnaires (that are standardized) are framed. We used a novel approach to determine "universal ranking order" and use that to determine weights that will reflect preferences. We used weighted normalized sums in assessing patient satisfaction. The primary virtue of our approach is that it puts the burden on "patients" and "service providers", the two primary stakeholders in "patient satisfaction" study, The methodology that we have put forth needs extensive validation, for which we need "real-life" data, collected from patients on questionnaires. We are hoping to do that in future.

References

[1] M. Cox and D. Ellsworth, "Application-controlled demand paging for out-of-core visualization," in *Proc. The 8th IEEE Conference on Visualization*. Los Alamitos, California, USA: IEEE Computer Society Press, 1997, pp. 235–244.

[2] D. Laney, "3d data management: Controlling data volume, velocity and variety (pdf) - gartner," META Group Report, Stamford, CT, USA, 2001.

[3] J. Andrew-Perez, C. C. Y. Poon, R. D. Merrifield, S. T. C. Wong, and G. Yang, "Big data for health," *IEEE Journal of Biomedical and Health Informatics*, vol. 19(4), pp. 1193–1208, 2015.

[4] M. Viceconti, P. Hunter, and R. Hose, "Big data, big knowledge: Big data for personalized healthcare," *IEEE Journal of Biomedical and Health Informatics*, vol. 19(4), pp. 1209–1215, 2015.

[5] M. Cottle, W. Hoover, S. Kanwal, M. Kohn, T. Strome, and N. W. Treister, "Transforming health care through big data," Institute for Health Technology Transformation, Washington DC, 2013.

[6] K. Wan and V. Alagar, "Characteristics and classification of big data in health care sector," in *In Proc. of 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, Changshaw, China, August 2016, pp. 1439–1446.

[7] W. Raghupathi and V. Raghupathy, "Big data analytics in healthcare: promise and potential," *Health Informattion and Systems: BioMed Central*, vol. 4:5, pp. 79–105, 2013.

[8] P. James, "Protein identification in the post-genome era: the rapid rise of proteomics," *Quarterly Review of Biophysics*, vol. 30, pp. 279–331, April 1997.

[9] D. Bennett, "Growing pains for metabolomics," *The Scientist*, vol. 19, pp. 25–28, August 2005.

[10] K. Wan and V. Alagar, "Dependable healthcare service automation: A holistic approach," in *Advanced Intelligent Computing Theories and Applications: Proc. of 11th International Conference, ICIC 2015, Part III*, Fuzhou, August 2015, pp. 766–777.

[11] D. P. Kessler and D. Mylod, "Does patient satisfaction affect patient loyalty?" *International Journal of Healthcare Quality Assurance*, vol. 24(4), pp. 266–273, 2011.

[12] N. Nguyen and V. Alagar, "Context aware mobile health care framework," Research Report, Department of Computer Science, Concordia University, May 2014.

[13] V. Zeithaml and A. Parasuraman, *Delivering Quality Service: Balancing Customer Perceptions and Expectations*. New York: Free Press, 1990.

[14] V. K. Iyer, "Exploring the factors affecting employee satisfaction and its influence on patient satisfaction in a community healthcare system in southwest wisconsin," MBA Thesis Report: Faculty of the Dhal School of Business, Viterbo University, Tech. Rep., 2015.

[15] S. Joachim, G. Katja, C. Stephen, B. Bjoem, R. Bernd, and W. Michael, "Is the job satisfaction of primary care team members associated with patient satisfaction?" *Medical Journal Quarterly Safety*, vol. 20, pp. 508–514, 2011.

[16] J. Gibbons, "Employee engagement: A review of current work and its implications," in *the Conference Board of Canada (E-0010-06-RR)*, 2006, pp. 1–18.

[17] W. Kahn, "Psychological conditions of personal engagement and disengagement at work," *The Academy of Management Journal*, vol. 33(4), pp. 692–724, 1990.

[18] I. Janicijevic, K. Seke, K. Djokovic, and J. Filipovic, "Healthcare workers satisfaction and patient satisfaction - where is the linkage?" *Hippokratia*, vol. 17(2), pp. 157–162, 2013.

[19] M. Ott and H. van Dijk, "Effect of hrm on client ei satisfaction in nursing care for the elderly," *Employee Relations*, vol. 27(4), pp. 413–424, 2005.

[20] J. Harter, F. Schmidt, S. Agrawal, and S. Plowman, "The relationship between engagement at work and organizational outcomes," http://www.employeeengagement.com/wp-content/uploads/2013/04/2012-Q12-Meta-analysis-research-paper.pdf, Gallup, Selection Research, Inc. and SRI, Tech. Rep., October 2013.

[21] V. Alagar and K. Wan, "Integrating trust and economic theories with knowledge science for dependable service automation," in *International Workshop on Security and Trust Management (STM 2014): Security and Trust Management*. Springer-Verlag, LNCS 8743, 2014, pp. 1–16.

[22] R. Oliver, "A cognitive model of the antecedents and consequences of satisfaction decisions," *Journal of Market Research*, vol. 17, pp. 460–469, 1980.

[23] R. I. Oliver, *Satisfaction: A behavioral perspective on the consumer*. Boston: McGraw-Hill, 1997.

[24] J. Lee and K. See, "Trust in automation: Designing for appropriate reliance," *Human Factors*, vol. 46(1), pp. 50–80, 2004.

[25] M. Fritz, T. Hausen, G. Schefer, and M. Canavari, "Trust and electronic commerce in the agrifood sector: a trust model and experimental experience," in *The* XIth International Congress of the EAAE (European Association of Agricultural Economists)*: The Future of Rural Europe in the Global Agri-Food System*, Copenhagen, August 2005, pp. 24–27.

[26] M. Hojat, D. Z. Louis, K. Maxwell, F. W. Markham, R. C. Wender, and J. S. Gonnella, "A brief instrument to measure patient's overall satisfaction with primary care physicians," *Family Medicine*, vol. 43(6), pp. 412–417, 2011.

[27] E. A. Platonova, K. N. Kennedy, and R. M. Shewchuk, "Understanding patient satisfaction, trust, and loyalty to primary care physicians," *Primary Care Research Review*, vol. 9, pp. 1–17, 2008.

[28] G. Cowan, *Statistical Data Analysis*. Clarendon Press, 1998.

[29] L. Alrubaiee and F. Alkaa'ida, "The mediating effect of patient satisfaction in the patients' perceptions of health quality - patient trust relationship," *International Journal of Marketing Studies*, vol. 3(1), pp. 103–127, 2011.

[30] G. Gabay and D. Moore, "Antecedents of patient trust in health-care insurers," *Serice Marketing Quarterly*, vol. 36(1), pp. 77–93, 2015.

[31] Wikipedia, "Schulze method," http://wiki.electorama.com/wiki/Schulze_method, 2017.

[32] S. Stahl and P. Johnson, *Understanding Modern Mathematics*. Sudbury: Jones and Bartlett Publishers, 2006.

[33] D. Mahoney, "Understanding the needs of the health care consumer," http://www.pressganey.com/resources, Resources & Research, Tech. Rep., 2017.

[34] AHRQ, "Cahps surveys and guidance," https://www.ahrq.gov/cahps/surveys-guidance/index.html, Agency for Healthcare Research and Quality, Tech. Rep., 2017.